

DETEKSI BIAS BUTIR DAN BIAS TES PADA SOAL MATEMATIKA UJIAN SEKOLAH DI SMP NEGERI 20 BULUKUMBA BERDASARKAN TEORI RESPON BUTIR

Oleh : Muliana

Penelitian dan Evaluasi Pendidikan
PPs, Universitas Negeri Makassar, Makassar
mulianamat01@gmail.com

Abstract

The objectives of the research were to discover (i) the characteristics of mathematics school exam at SMPN 20 Bulukumba reviewed from item biased (differential item functioning) based on items response theory; (ii) the characteristics of mathematics school exam at SMPN 20 Bulukumba reviewed from test biased (differential test functioning) based on items response theory.

The research was explorative research to discover items and test biased based on items response theory. The subjects of the research were questions and all of the students' answer sheets of class IX at SMPN 20 Bulukumba of 2013/2014 academic year. The subject of the research was taken by using purposive sampling technique and obtained 286 responses of the students which consisted of 147 male and 149 female students' responses. The data of the research was collected by using documentation technique by quoting the response of mathematics school exam participants at SMPN 20 Bulukumba of 2013/2014 academic year. While the data analysis technique of the research was quantitative by using items response theory approach model 2P with Lord's chi-square DIF method. In this case, the data analysis of the research covered two main elements, namely the detection of Differential Item Functioning (DIF) and Differential Test Functioning (DTF).

The result of the research showed that from 40 question items which were analyzed theoretically by using items response theory reviewed from items biased (DIF) infected DIF gender was 10 items, and infected DIF location (area) was 13 items. Meanwhile, reviewing from test biased (DTF) was benefitted for female and least profitable to city people.

Keywords : Differential Item Functioning (DIF), Differential Test Functioning (DTF), Item Response Theory.

I. PENDAHULUAN

Ujian Sekolah merupakan salah satu proses pengukuran terhadap hasil belajar tingkat sekolah memiliki tujuan dan fungsi yang sangat penting dalam bidang pendidikan. Hasil ujian sekolah yang diberikan dalam bentuk nilai ujian sekolah (UAS) sangat diharapkan dapat digunakan untuk melihat gambaran kemampuan peserta didik secara murni serta mampu berfungsi sebagai indikator keberhasilan proses pendidikan. Mengingat pentingnya nilai ujian sekolah tersebut, maka soal yang disusun harus benar-benar dapat mengukur apa yang seharusnya diukur. Hasil pengukuran diharapkan akurat sehingga dapat dipercaya, dan

mencerminkan kemampuan peserta didik yang sebenarnya. Artinya, perbedaan skor yang diperoleh peserta didik dengan peserta didik lainnya semata-mata hanya karena perbedaan kemampuan di antara mereka, bukan disebabkan oleh faktor lain, seperti ada bias pada butir tes yang mereka kerjakan.

Dalam pengukuran pendidikan istilah bias pada suatu butir dikenal sebagai differential item functioning (DIF) dan differential Test Functioning (DTF) Holland & Thayer (dalam Camilli & Shepard, 1994 : 62). Berbagai teknik atau metode pendeteksian DIF telah banyak ditemukan dan digunakan selain keberfungsian butir diferensial, untuk mengindikasikan suatu tes adil bagi semua golongan atau tidak, dapat

digunakan keberfungsian tes diferensial atau differential Test Functioning (DTF). Keberfungsian tes diferensial ini merupakan hal yang penting bagi suatu soal karena DTF dapat menentukan apakah suatu tes adil bagi semua kelompok peserta tes atau tidak.

Wand & Brown (dalam Anas, 2012 : 2) yang menyatakan bahwa “ evaluasi pendidikan adalah kegiatan atau proses penentuan nilai pendidikan, sehingga dapat diketahui mutu atau hasil-hasinya”. Sedangkan menurut Grounlund (dalam Djaali & Muljono, 2008 : 1) menyatakan bahwa “ suatu proses yang sistematis untuk menentukan atau membuat keputusan, sampai sejauh mana tujuan program telah tercapai”.

Mengenai evaluasi terhadap aspek tes hasil belajar matematika dalam pendidikan khususnya pembelajaran matematika, akan memerlukan instrumen yang diantaranya berupa tes. Tes merupakan sejumlah pertanyaan yang memiliki jawaban yang benar atau salah. “Tes juga diartikan sejumlah pertanyaan yang membutuhkan jawaban, atau sejumlah pertanyaan yang harus diberikan tanggapan dengan tujuan mengukur tingkat kemampuan seseorang atau mengungkap aspek tertentu dari yang dikenai tes” (Mansyur dkk, 2009 : 21).

Teori respon butir adalah model matematika yang berhubungan dengan kemungkinan kemampuan peserta ujian merespon suatu item tertentu secara keseluruhan (Camilli & Shepard, 1994 : 46). Teori ini secara umum memiliki ciri-ciri sebagai berikut: 1) Karakteristik butir tidak tergantung pada peserta ujian, 2) skor yang digambarkan peserta ujian tidak tergantung pada tes, 3) merupakan model yang lebih menekankan pada tingkat butir dari pada tingkat tes, 4) merupakan model yang tidak mensyaratkan secara ketat tes paralel untuk menaksir reliabilitas, dan 5) merupakan model yang menguraikan sebuah ukuran tiap skor kemampuan yakni ada hubungan fungsional antara peserta ujian terhadap tingkat kemampuan yang dimiliki.

Model matematik dalam IRT mengatakan bahwa probabilitas subjek untuk menjawab suatu butir dengan benar tergantung pada kemampuan subjek dan karakteristik butir. Untuk itu dalam IRT ada asumsi pendukung

yang secara tidak langsung dapat diukur dan dibuktikan adanya. Asumsi yang melandasi teori respon butir adalah unidimensi, independensi lokal, dan fungsi karakteristik butir atau kurva karakteristik butir. Unidimensi artinya bahwa dimensi karakter peserta yang diukur oleh suatu tes itu tunggal. Dalam konteks tes kemampuan, dimensionalitas disebut sebagai banyaknya kemampuan yang diukur oleh tes atau oleh kumpulan item (Aswar & Ridho, 2012 : 2).

Parameter pada teori respon butir adalah tingkat kesulitan soal disimbol b , daya beda disimbol a , dan tebakan disimbol c . “informasi pola respon terhadap suatu tes atau instrumen lain yang digunakan untuk mengestimasi besarnya kemampuan seseorang. Estimasi kemampuan dan parameter butir menggunakan maksimum Likelihood (MLE) dan Bayesian” (Mardapi, 2012 : 204).

Selain ketiga karakteristik butir yang perlu diperhatikan dalam teori respon butir, sebenarnya masih ada satu hal yang perlu diperhatikan yaitu fungsi informasi butir (item information function). Fungsi informasi butir ini merupakan suatu metode untuk menjelaskan kekuatan suatu butir pada soal yang menyatakan suatu sumbangan butir soal dalam mengungkap kemampuan laten (laten trait) yang diukur dengan tes. Dengan diketahui fungsi informasi butir maka dapat diketahui butir mana yang cocok dengan model sehingga membantu dalam seleksi butir soal.

Harapan pemerintah yang tertuang dalam Permendiknas nomor 20 tahun 2007 tentang standar penilaian menyatakan bahwa salah satu prinsip penilaian dalam pendidikan adalah adanya unsur keadilan dan diadopsi dalam Permendiknas nomor 66 tahun 2013 yang menyatakan bahwa salah satu prinsip penilaian adalah sesuai dengan konteks sosial budaya. Dengan demikian, makna dari adanya unsur keadilan dan sesuai dengan konteks sosial budaya ini adalah salah satu persyaratan instrumen itu harus sah dan handal serta mantap dan adil sehingga tidak merugikan salah satu kelompok tertentu baik dari segi budaya, wilayah, suku, ras dan gender.

Berdasarkan UUD 1945, setiap warga Negara berhak mendapatkan pendidikan (pasal 31). Ini berarti setiap warga Negara baik perempuan maupun laki-laki memperoleh pendidikan. Sebagai akibatnya, pada tes akhir ujian sekolah diperlukan suatu tes yang tidak memihak pada kelompok peserta didik tertentu, baik dari segi suku agama, jenis kelamin dan sebagainya. Untuk menjamin diperolehnya gambaran peserta didik yang berkualitas diperlukan suatu tes atau soal yang memiliki fungsi informasi yang tinggi.

Prosedur dalam mendeteksi bias butir yang digunakan akan menentukan apakah butir soal yang diberikan akan memberikan informasi yang valid. Untuk itu, maka digunakan beberapa cara untuk mendeteksi bias butir dan bias tes agar perangkat tes adil untuk semua peserta tes.

Pendeteksian bias tes (DTF) dapat dilakukan berdasarkan teori respon butir yaitu dengan melihat grafik nilai peluangnya. Adapun untuk pendeteksian bias butir (DIF) pada suatu tes berdasarkan teori respon butir dapat dilakukan dengan berbagai pendekatan yaitu; (1) uji perbedaan parameter kesukaran butir, (2) metode penyimpangan butir, (3) uji kai kuadrat Lord, (4) distribusi sampling empirik untuk indeks DIF, dan (5) perbandingan model teori respon butir (Camilli & Shepard, 1994 : 70-74). sedangkan, "Pendekatan perbandingan model teori respon butir dibagi menjadi empat kategori utama yaitu : a) pengantar umum tentang likelihood, b) rasio likelihood berdasarkan statistik inferensial, c) pendekatan model perbandingan aplikasi analisis DIF, dan d) teori respon butir 3 parameter" (Camilli dan Shepard, 1994 : 74-80).

Perbedaan parameter menyebabkan DIF terjadi dalam dua kategori umum yaitu: (1) DIF konsisten atau uniform yang terjadi jika kurva karakteristik butir berbeda dan tidak langsung berpotongan atau bersilangan, dan (2) DIF tidak konsisten atau non uniform yang terjadi jika kurva karakteristik butir berbeda tapi berpotongan pada suatu titik pada skala θ (Camilli & Shepard, 1994 : 64). Untuk mengetahui kelompok mana yang lebih diuntungkan, dapat dilihat secara langsung dari grafik nilai peluangnya.

Menurut Permendiknas Nomor 20 Tahun 2007 tentang Standar Penilaian oleh Pendidikan, Ujian Akhir Sekolah (UAS) adalah kegiatan yang dilakukan oleh pendidik di akhir untuk mengukur pencapaian kompetensi peserta didik di akhir pada satuan pendidikan.

Menurut Naga (1992 : 380) mengemukakan bahwa " bank soal secara sederhana dapat didefinisikan sebagai kumpulan butir tes". Akan tetapi bank soal bukan sekedar kumpulan soal-soal saja. Butir-butir soal yang ada di dalamnya merupakan butir yang telah diseleksi melalui prosedur atau informasi tertentu.

Adapun kriteria kuaalitas butir yang baik menurut teori respon butir mengacu pada masing-masing parameter butir. Kriteria untuk tingkat kesukarang (b), butir-butir soal yang memiliki nilai lebih dari 2 atau $b > 2$ adalah butir soal yang dianggap terlalu sukar (Hambleton, dkk, 1991). Butir yang terlalu sukar tidak dapat melakukan fungsi ukurnya dengan baik, karena peserta tes akan cenderung menjawab dengan menggunakan tebakan. Harga parameter tingkat kesukaran yang baik berkisar antara 0 sampai 2. Butir soal yang memiliki harga parameter lebih kecil dari -2 adalah butir soal yang terlalu mudah dan harus direvisi.

Sementara untuk criteria daya beda soal (a), Hambleton dkk (1991) menjelaskan apabila butir soal tersebut harus direvisi atau dibuang, sedangkan daya pembeda > 2 jarang terjadi. Sehingga daya beda soal berkisar antara 0 sampai 2 menunjukkan bahwa butir soal tersebut dapat membedakan antara peserta tes yang berkemampuan tinggi dan peserta tes yang berkemampuan rendah.

Untuk mendeteksi keberfungsian butir diferensial (bias butir) berdasarkan Teori Respon Butir model logistik 2 parameter menggunakan Lord's chi-square method sedangkan untuk mendeteksi keberfungsian tes diferensial berdasarkan teori respon butir model logistic 2 parameter dengan melihat grafik nilai peluang pada masing-masing butir tergantung kelompok mana yang diuntungkan dan kelompok mana yang dirugikan. dimana, dalam penelitian ini mendeteksi dua jenis DIF yaitu (1) DIF berdasarkan perbedaan gender, (2) DIF berdasarkan perbedaan lokasi.

Berdasarkan latar belakang masalah di atas, maka permasalahan dalam penelitian ini dirumuskan sebagai berikut.

1. Bagaimana karakteristik soal matematika Ujian Sekolah di SMP Negeri 20 Bulukumba tahun pelajaran 2013/2014 ditinjau dari bias butir (keberfungsian butir diferensial) berdasarkan teori respon butir?
2. Bagaimana karakteristik butir soal matematika ujian sekolah (UAS) di SMP Negeri 20 Bulukumba tahun pelajaran 2013/2014 ditinjau dari bias tes (Keberfungsian tes diferensial) berdasarkan teori respon butir?

Berdasarkan rumusan masalah maka tujuan penelitian ini adalah untuk mengetahui.

1. Karakteristik soal matematika Ujian Sekolah di SMP Negeri 20 Bulukumba ditinjau dari bias butir (keberfungsian butir diferensial) berdasarkan teori respons butir;
2. Karakteristik butir soal matematika ujian sekolah di SMP Negeri 20 Bulukumba tahun ajaran 2013/2014 ditinjau dari bias tes (Keberfungsian tes diferensial) berdasarkan teori respon butir.

II. METODE

Penelitian ini bersifat eksploratif, yakni akan mengungkap bias butir (keberfungsian butir diferensial) dan bias tes (keberfungsian tes diferensial) berdasarkan teori respon butir

Penelitian ini dilaksanakan di SMP Negeri 20 Bulukumba. Adapun alasan saya memilih SMP Negeri 20 Bulukumba dengan pertimbangan terpenuhinya asumsi teori respon butir. Dengan menggunakan kriteria sample size mulai dari 50, 200, dan 1000 dengan panjang tes 10, 20, 80", Hambleton & Cook (dalam Foley, 2010 : 15-16). Penelitian ini dilaksanakan 3 bulan

Data dalam penelitian ini adalah respon peserta ujian Sekolah pada soal mata pelajaran Matematika di SMP Negeri 20 Bulukumba tahun pelajaran 2013/2014. Sumber data berupa lembar jawaban peserta didik yang telah

didokumentasi. Soal milik sekolah SMP Negeri 20 Bulukumba yang dibuat oleh tim guru mata pelajaran matematika yang terdiri dari 3 orang yaitu: Rusman, Nurfitriani, dan Rosmaniar.

Subjek dalam penelitian ini adalah soal dan seluruh lembar jawaban peserta didik kelas IX SMP Negeri 20 Bulukumba tahun pelajaran 2013/2014. Pemilihan subjek dalam penelitian ini dengan menggunakan teknik purposif sampling sebanyak 286 respon peserta didik yang terdiri atas 137 respon peserta didik laki-laki dan 149 respon peserta didik perempuan.

Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan teknik dokumentasi, yaitu dengan mengutip respon peserta Ujian Sekolah pada mata pelajaran Matematika di SMP Negeri 20 Bulukumba tahun pelajaran 2013/2014.

Teknik analisis data dalam penelitian ini adalah analisis data secara kuantitatif dengan menggunakan pendekatan teori respon butir. Dalam hal ini kegiatan analisis data dalam penelitian meliputi dua hal pokok yaitu sebagai berikut.

Pendeteksian DIF dengan menggunakan pendekatan Lord's chi-squared DIF method dengan kategori 2P, dengan rumusnya sebagai berikut:

$$P_i(\theta) = \frac{1}{1 + e^{D a_i (\theta - b_i)}} \quad (\text{Sudaryono, 2012 : 200}).$$

Keterangan rumus:

$P_i(\theta)$ = peluang subjek yang memiliki kemampuan θ untuk menjawab aitem I dengan benar

b_i = parameter tingkat kesukaran butir

a_i = parameter daya pembeda butir

c_i = parameter unsur tebakan

e = angka transcendental yang bernilai 2,178

D = faktor penskalaan yang bernilai 1,7

Langkah-langkah dalam mendeteksi DIF Lord berdasarkan gender sebagai berikut.

1. Respon peserta ujian diinput dengan bantuan program Excel kemudian dilanjutkan keprogram SPSS;
2. Dari program SPSS dimasukkan keprogram R dengan mengikuti langkah berikut:

- a. Instal packages → Indonesia Jakarta
→ diFR sehingga muncul (Loading required package: lme4, Loading required package: Matrix, Loading required package: Rcpp, Loading required package: ltm, Loading required package: MASS, Loading required package: msm, Loading required package: polycor, Loading required package: mvtnorm, Loading required package: sfsmisc, Warning messages: (1) package 'diFR' was built under R version 3.1.3, (2) package 'lme4' was built under R version 3.1.3, (3) package 'Rcpp' was built under R version 3.1.3, (4) package 'ltm' was built under R version 3.1.3.
- b. library(diFR);
- c. library(foreign);
- d. muliana <- read.spss("D://DIF/KK.sav", to.data.frame=TRUE)
- e. muliana;
- f. keempat langkah di atas (a,b,c,dan d kemudian (run line or selection Ctrl+R);
3. ??diFR pilih diFR-diFR package *Collection of methods to detect dichotomous differential item functioning (DIF) in psychometric*. Sehingga muncul pilihan beberapa metode DIF ;
4. Memilih model IRT yang akan digunakan mengestimasi daya beda butir dan tingkat kesukaran butir dengan mengikuti perintah *itemPar2PL(muliana [,1:40])*
5. Memilih metode yang akan digunakan dalam mendeteksi DIF (*Lord's chi-squared DIF method*);
6. Mengikuti perintah *DIFLord* (*difLord(muliana, group="Gender", focal.name=1, model="2PL", purify=TRUE)*);
7. Dengan mengikuti langkah 1-6 maka muncullah hasil DIF Gender yang ditandai dengan tanda *;
8. Untuk melihat terjadi DIF konsisten (*uniform*) atau DIF tidak konsisten (*non uniform*) dengan cara melihat grafik nilai peluangnya. DIF konsisten terjadi apabila keuntungan dalam satu kelompok terhadap kelompok lainnya terjadi pada setiap level kemampuan atau grafik tidak berpotongan pada satu titik pada skala kemampuan

sedangkan DIF tidak konsisten jika keuntungan tidak terjadi pada setiap level kemampuan atau jika grafik berpotongan pada satu titik pada skala kemampuan.

Langkah-langkah pendeteksian DIF berdasarkan lokasi dalam metode DIF Lord's diuraikan sebagai berikut.

1. Respon peserta ujian diinput dengan bantuan program Excel kemudian dilanjutkan keprogram SPSS;
2. Dari program SPSS dimasukkan ke program R dengan mengikuti langkah berikut:
 - a. Instal packages → Indonesia Jakarta
→ diFR sehingga muncul (Loading required package: lme4, Loading required package: Matrix, Loading required package: Rcpp, Loading required package: ltm, Loading required package: MASS, Loading required package: msm, Loading required package: polycor, Loading required package: mvtnorm, Loading required package: sfsmisc, Warning messages: (1) package 'diFR' was built under R version 3.1.3, (2) package 'lme4' was built under R version 3.1.3, (3) package 'Rcpp' was built under R version 3.1.3, (4) package 'ltm' was built under R version 3.1.3.
 - b. library(diFR);
 - c. library(foreign);
 - d. muliana <- read.spss("D://DIF/KK.sav", to.data.frame=TRUE);
 - e. muliana;
 - f. keempat langkah di atas (a,b,c,dan d kemudian (run line or selection Ctrl+R) sehingga muncul respon variabel mulai butir 1-40 serta hasil estimasi gender yang disimbol dikotomi;
3. ??diFR pilih diFR-diFR package *Collection of methods to detect dichotomous differential item functioning (DIF) in psychometric*. Sehingga muncul pilihan beberapa metode DIF ;
4. Sebelum mendeteksi DIF maka terlebih dahulu memilih model IRT 2P yang akan digunakan mengestimasi daya beda butir dan tingkat kesukaran butir dengan perintahnya adalah *itemPar2PL(muliana[,1:40])*;

5. Memilih metode yang akan digunakan dalam mendeteksi *DIF* (*Lord's chi-squared DIF method*);
6. Mengikuti perintah *DIFLord* (*difLord(muliana, group="Region", focal.name=1, model="2PL", purify=TRUE)*);
7. Dengan mengikuti langkah 1-5 maka muncullah hasil *DIF* lokasi yang ditandai dengan tanda *;
8. Untuk melihat terjadi *DIF* konsisten (*uniform*) atau *DIF* tidak konsisten (*non uniform*) dengan cara melihat grafik nilai peluangnya. *DIF* konsisten terjadi apabila keuntungan dalam satu kelompok terhadap kelompok lainnya terjadi pada setiap level kemampuan atau grafik tidak berpotongan pada satu titik pada skala kemampuan sedangkan *DIF* tidak konsisten jika keuntungan tidak terjadi pada setiap level kemampuan atau jika grafik berpotongan pada satu titik pada skala kemampuan.

III. HASIL DAN PEMBAHASAN

Hasil analisis validasi butir pengetahuan (tes hasil belajar matematika) diawali dengan memberikan hasil rancangan instrument pada validator. Validasi isi oleh pakar melibatkan dua orang pakar pendidikan matematika, yang keduanya merupakan dosen FMIPA UNM. Hasil penilaian dan analisis validasi pakar dapat dilihat pada gambar berikut.

Dari penilaian yang diberikan oleh kedua validator diatas menunjukkan satu butir kurang relevan (sel A) yaitu butir 14, pada butir 9 validator pertama memberikan penilain relevan tetapi validator kedua memandangnya kurang relevan maka butir ini dimasukkan ke dalam sel B, pada butir 6 validator pertama memberikan penilaian tidak relevan tetapi validator kedua memandangnya kurang relevan maka butir ini dimasukkan ke dalam sel C dan 37 butir termasuk sangat relevan (sel D) sehingga dapat dihitung tingkat validasinya berdasarkan rumus konsistensi internal model Gregory sebagai berikut. Koefisien Konsistensi Internal

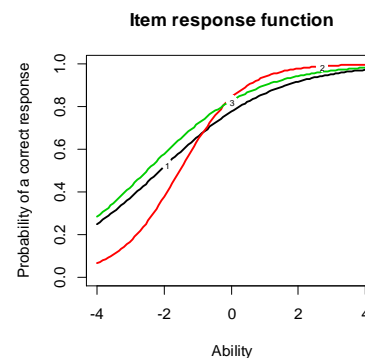
$$= \frac{D}{(A+B+C+d)}$$

$$= \frac{37}{(1+1+1+37)}$$

$$= 0,93$$

Jadi dapat disimpulkan bahwa validitas yang diperoleh yaitu 0,93 atau $V = 93\%$. Hal ini berarti bahwa hasil penilaian kedua validator $>75\%$ sehingga memiliki kriteria “relevansi yang kuat” (Ruslan, 2009 : 19)

I. Karakteristik Soal UAS SLTP Negeri 20 Bulukumba pada Mata Pelajaran Matematika Tahun Pelajaran 2013/2014 Ditinjau dari Daya Pembeda Butir

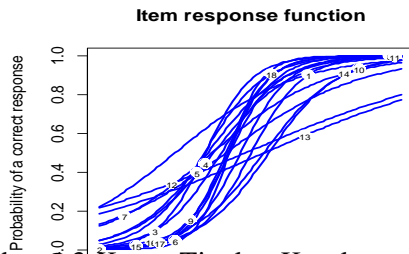


Gambar 4.2 Kurva Karakteristik Butir Model Logistik 2P Butir Soal Ujian Sekolah SMPN 20 Bulukumba Tahun Pelajaran 2013/2014

Berdasarkan hasil pengolahan data pada tabel di atas dapat ditarik kesimpulan bahwa dari 40 butir soal yang dianalisis berdasarkan teori respon butir model 2P diperoleh 7,5% butir soal berkategori sangat baik dan 30% serta 62,5% berkategori buruk.

II. Karakteristik Soal UAS SMP Negeri 20 Bulukumba pada Mata Pelajaran Matematika Tahun Pelajaran 2013/2014 Ditinjau dari Tingkat Kesukaran Butir

Berdasarkan hasil pengolahan data mengenai tingkat kesukaran soal-soal pilihan ganda seperti terlihat pada tabel 4.3 di atas, diperoleh bahwa sebanyak 3 butir atau 7,5 % tingkat kesukaran tes berada pada kategori mudah, 3 butir atau 7,5 % berada pada kategori sedang, 12 butir atau 30 % berkategori sangat mudah, 5 butir atau 12,5 % berkategori sukar, dan 17 butir atau 42,5 % berkategori sangat sukar.

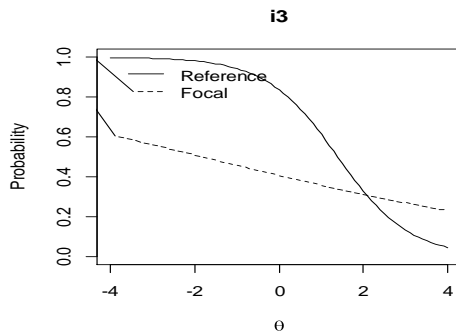


Gambar 4.3 Kurva Tingkat Kesukaran Butir (b) Soal Ujian Sekolah Tahun Pelajaran 2013/2014 SMPN 20 Bulukumba Berdasarkan Teori Respon Butir Model logistik 2P

IV. Deskripsi Hasil Penelitian Berdasarkan Keberfungsian Butir Deferensial (DIF)

a. Bias Butir berdasarkan Gender

1. Grafik Bias Butir 3



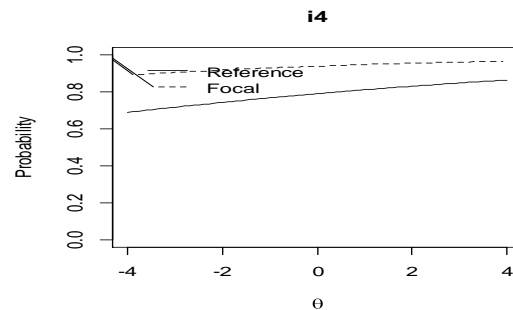
Gambar 4.5 Kurva DIF pada Soal Ujian Sekolah Tahun Pelajaran 2013/2014 SMPN 20 Bulukumba Berdasarkan Teori Respon Butir Model Logistik 2P

Gambar di atas menunjukkan bahwa pada level kemampuan 2 grafik tersebut berpotongan, ini menandakan bahwa pada level kemampuan tersebut sebagaimana teori yang dijelaskan pada BAB. II bahwa DIF konsisten terjadi jika pada grafik DIF tidak berpotongan dan (DIF) tidak konsisten terjadi jika grafik tersebut berpotongan pada salah satu titik dan tidak terjadi pada setiap level atau skala kemampuan (ability). Oleh karena itu, pada gambar di atas menunjukkan DIF tidak konsisten (non uniform). dikategorikan terjangkit bias berdasarkan gender.

b. Bias berdasarkan lokasi (wilayah)

Pada bias lokasi mengindikasikan kelompok peserta ujian yang berasal dari kota sebagai kelompok fokus (fokal) dan peserta ujian yang berasal dari desa sebagai kelompok referensi (reference). Untuk lebih jelasnya perhatikan masing-masing gambar bias lokasi berikut.

1. Grafik Bias Lokasi Butir 4



Gambar 4.15 Kurva DIF pada Soal Ujian Sekolah Tahun Pelajaran 2013/2014 SMPN 20 Bulukumba Berdasarkan Teori Respon Butir Model Logistik 2P

Gambar di atas menunjukkan bahwa grafik tidak berpotongan pada tiap level kemampuan. ini menandakan bahwa DIF terjadi pada tiap level kemampuan. Oleh karena itu, pada gambar di atas menunjukkan DIF konsisten (uniform).

Berdasarkan hasil analisis data menurut Teori Respon Butir model logistik 2 parameter dengan menggunakan Lord's chi-square method versi 386.3.1.2 pada Soal Ujian Sekolah SMP Negeri 20 Bulukumba Tahun Pelajaran 2013/2014 menunjukkan bahwa dari 40 butir soal, 3 butir soal dengan kategori sangat baik, 12 butir kategori baik, dan 25 butir berkategori buruk. Dengan demikian, 15 butir soal dapat dipilih dalam kegiatan pengembangan bank soal yaitu butir 1, 2, 5, 6, 7, 9, 11, 12, 16, 22, 23, 28, 29. Namun ditinjau dari segi keberfungsian butir diferensial dan keberfungsian tes diferensial ternyata dari 40 butir soal yang dideteksi bias butir berdasarkan gender dan bias berdasarkan lokasi (wilayah). Hal ini dapat dilihat berdasarkan tabel berikut.

Tabel 4.3 Hasil Deteksi Bias Gender pada Soal Matematika Ujian Sekolah SMPN 20 Bulukumba Tahun Pelajaran 2013/2014 Berdasarkan Teori Respon Butir Model Logistik 2P dengan menggunakan Metode Lord Chi-Square

Bias Gender	Berdasarkan Grafik Nilai Peluang	
	Bias konsisten	Bias tidak Konsisten
10	6	4

Berdasarkan tabel sehingga dapat diuraikan bahwa dari 40 butir soal terdapat 10 butir (25%) terindikasi mengandung bias gender sehingga jika dilihat dari grafik nilai peluangnya ternyata 6 butir memuat bias konsisten dan 4 butir memuat bias tidak konsisten yang mana butir-butir tersebut cenderung menguntungkan kelompok perempuan dan merugikan kelompok laki-laki. Selain soal tersebut mengandung bias gender juga terdeteksi mengandung bias lokasi, hal ini dapat dilihat pada tabel berikut.

Tabel 4.4 Hasil Deteksi Bias Lokasi pada Soal Matematika Ujian Sekolah SMPN 20 Bulukumba Tahun Pelajaran 2013/2014 Berdasarkan Teori Respon Butir Model Logistic 2P dengan menggunakan Metode Lord Chi-Square

Bias Lokasi	Berdasarkan Grafik Nilai Peluang	
	Bias konsisten	Bias tidak Konsisten
13	6	7

Tabel di atas menunjukkan bahwa dari 40 butir soal terdapat 13 butir (32,5%) terindikasi mengandung bias gender sehingga jika dilihat dari grafik nilai peluangnya ternyata 6 butir memuat bias konsisten dan 7 butir memuat bias tidak konsisten yang mana butir-butir tersebut cenderung menguntungkan kelompok peserta tes yang berlokasi di kota dan merugikan kelompok peserta tes yang berlokasi di desa.

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis data dan pembahasan penelitian yang telah dikemukakan dan dihubungkan dengan rumusan masalah, ada beberapa hal yang berkaitan dengan deteksi bias butir (*DIF*) dan bias tes (*DTF*) pada Soal Matematika Ujian Sekolah SMP Negeri 20 Bulukumba yaitu.

1. Karakteristik soal matematika Ujian Sekolah di SMP Negeri 20 Bulukumba terdeteksi 10 butir soal terindikasi bias butir (*DIF*) berdasarkan gender dan 13 butir terindikasi bias berdasarkan lokasi (wilayah).
2. Karakteristik soal matematika Ujian Sekolah di SMP Negeri 20 Bulukumba terdeteksi memuat bias tes (*DTF*) 6 butir tidak konsisten dan bias tes konsisten sebanyak 6 butir berdasarkan gender yang cenderung menguntungkan perempuan sedangkan berdasarkan lokasi memuat bias tes konsisten 6 butir dan 7 butir memuat bias tes tidak konsisten yang cenderung menguntungkan orang kota.

Berdasarkan hasil yang diperoleh dalam penelitian, terdapat beberapa hal yang dianggap perlu untuk diperhatikan terkait dengan deteksi bias butir berdasarkan teori respon butir model 2P pada metode Lord's *chi-square statistic*.

1. Kepada penyusun soal matematika SMP baik dalam skala lokal maupun skala nasional sebaiknya menggunakan soal yang secara empirik terbukti berkualitas baik dan tidak mengandung *DIF* dan *DTF*;
2. Pembuktian secara empirik terhadap kualitas suatu butir perlu lebih dikembangkan dan disosialisasikan pada semua praktisi pendidikan misalnya dengan mengadakan pelatihan bagi guru-guru baik bidang studi matematika maupun bidang studi lainnya;
3. Untuk mendeteksi *DIF* dan *DTF* suatu tes dapat dilakukan berdasarkan teori respon butir;
4. Untuk lebih memperluas kajian tentang *DIF* dan *DTF* perlu adanya penelitian lebih lanjut dengan menggunakan metode lain
5. Untuk peneliti lebih lanjut seharusnya menggunakan instrumen observasi sebagai pendukung hasil penelitian

V. DAFTAR PUSTAKA

- Ali, Sidin & Khaeruddin, 2012. Evaluasi Pembelajaran. Badan penerbit UNM
- Azward, Saifuddin, 2014. Dasar-Dasar Psikometri. Pustaka Pelajar. Yogyakarta
- Azward, S. & Ridho, A. 2012. Abilitas Komposit Dalam Tes Potensi. (online), (<http://google.com>, PAPs UGM, Posting 21)
- Djaali & Muljono, Pudji, 2008. Pengukuran Dalam Bidang Pendidikan. Gramedia Widiasarana Indonesia.
- Djemari, Kumaidi, & Badrun, 2011. Pengembangan Instrumen Pengukur Hasil Belajar Nirbias dan Terskala Baku. *Jurnal Penelitian dan Evaluasi Pendidikan*, (online), jilid 15, No. 2, (journal.uny.ac.id, Diakses 11 Desember 2014).
- Foley, Brett P., 2010. Improving IRT Parameter Estimates with Small sampel sizes: evaluating The Efficacy of A New Data Augmentation Technique. *Journal Education and Human Science, College Of (CEHS)*. (<http://digitalcommons.unl.edu/cehsdiss>, Diakses 1 juli 2010)
- Hambleton, R. K., Swaminathan H., & Rogers, H. J., 1991. Fundamentals of Item Response Theory. Newbury Park, CA: Sage Publication Inc.
- Hamzah, Ali, 2014. Evaluasi Pembelajaran matematika. PT RajaGrafindo Persada. Jakarta.
- Kartowagiran, Badrun, 2005. Perbandingan Berbagai Metode Untuk Mendeteksi Bias Butir. *Jurnal Penelitian dan Evaluasi Pendidikan*, (online), (staff.uny.ac.id, Diakses 12 Desember 2014).
- Kementrian Pendidikan dan Kebudayaan. 2007. Peraturan Menteri Pendidikan dan Kebudayaan Nomor 20 Tahun 2007 tentang Penilaian. Jakarta: Depertemen Pendidikan dan Kebudayaan.
- Kementrian Pendidikan dan Kebudayaan. 2013. Peraturan Menteri Pendidikan dan Kebudayaan Nomor 3 Tahun 2013 tentang Kriteria Kelulusan Peserta didik. Jakarta: Depertemen Pendidikan dan Kebudayaan.
- Kementrian Pendidikan dan Kebudayaan. 2013. Peraturan Menteri Pendidikan dan Kebudayaan Nomor 66 Tahun 2013 tentang Penilaian. Jakarta: Depertemen Pendidikan dan Kebudayaan.
- Naga, Dali. S, 1992. Pengantar Teori Sekor pada Pengukuran. Pendidikan. Gunadarma. Jakarta
- Mardapi, D., 1991. Konsep Dasar Teori Respon Butir: Perkembangan dalam Bidang Pengukuran Pendidikan. Cakrawala Pendidikan.
- Mardapi, D., 2012. Pengukuran Penilaian & Evaluasi Pendidikan. Nuha Litera. yogyakarta
- Mansyur, Harun & Rasyid, 2009. Asesmen Pembelajaran di Sekolah. Multi Pressindo. Yogyakarta.
- Margono, G. 2013. Aplikasi Analisis Faktor Komfirmatori Untuk Menentukan Reliabilitas Multidimensi. *Statistika*, (online), Vol 13. Nomor 1 (Diakses Mei 2013).
- Rahayu, W. 2008. Metode *Lingking* dan Butir *False Positive* pada Pendeteksian *DIF* Berdasarkan Teori Respon Butir. *Jurnal Penelitian dan dan Evaluasi Pendidikan*, (online). Vol 14. Nomor 1 (wardani9164@yahoo.com, Diakses Tahun 2010).
- Rasyid, Harun, & Mansur, 2007. *Penilaian Hasil Belajar*. CV. Wacana Prima. Bandung.

- Retnawati, H. & Hidayati, Kana, 2006. Pendeteksian Bias Tes dan Butir Perangkat Soal Matematika Ujian Nasional SLTP Berdasarkan Teori Respon Butir, (online), (staff.uny.ac.id, Diakses 12 desember 2014)
- Retnawati, H. 2005. Keberfungsian Butir Diferensial pada Perangkat Tes Seleksi Masuk SLTP Mata Pelajaran Matematika. *Jurnal Penelitian dan dan Evaluasi Pendidikan*,(online). Nomor 6 (staff.uny.ac.id, Diakses 15 Mei Tahun 2005).
- Ruslan. 2009. “Validitas Isi” Buletin Pabbiritta No. 10 Tahun VISeptember 2009. Makassar : LPMP Sulawesi Selatan
- Sappaile, B. I. 2007. Konsep Instrumen Penelitian Pendidikan. *Jurnal Penelitian dan Kebudayaan*,(online). Vol 13. Nomor 66 (<http://basointang.files.wodpress.com>, Diakses 15 Desember 2014).
- Sudaryono. 2012. *Dasar-dasar Evaluasi Pembelajaran*. Graha Ilmu. Tangerang
- Sudaryono. 2012. Kajian Metode Deteksi Differential Item Functioning (DIF) Butir Soal Ujian Nasional Dengan Teori Tes Klasik. *Jurnal Pendidikan dan Kebudayaan*. (Online), Vol. 18. Nomor 2 (<http://sippendidikan.kemdikbud.go>. Id, Diakses 11 Desember 2014).
- Sudiono, Anas, 2012. Pengantar Evaluasi Pendidikan. PT RajaGrafindo Persada. Jakarta.
- Suherman, E, 2003. *Strategi Pembelajaran Matematika Kontenporer*. FMIPA UPI. Bandung
- Tim Sisjian, 1997. *Bank Soal*. Pusat Penelitian dan Pengembangan Sistem Pengujian.
- Tiro, Arif, Muhammad. 2010. *Cara Efektif Belajar Matematika*. Adira Publisher. Makassar.
- Widhiarso, W., & Retnowati, S., 2011. Investigasi Bias Butir Jender dalam Pengukuran Depresi Melalui *Chidren's Depression Inventory (CDI)*. *Jurnal Penelitian Psikologi*. (online), Volume 02, No. 01, (journal.uny.ac.id, Diakses 11 Desember 2014).
- Widoyoko, Putro, Eko, 2014. *Teknik Penyusunan Instrumen*. Pustaka Pelajar. Yogyakarta.
- Zumbo, B., D., 1999. *A Handbook on the Theory and Methods of Diffrential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, On: Direktorat of Human Resources Research and Evaluation, Departemen of Nationalitio, Departement of National Defense.